

Distributed Uplink Resource Allocation in Cognitive Radio Networks – Part I: Equilibria and Algorithms for Power Allocation

Mingyi Hong, Alfredo Garcia and Stephen G. Wilson

Abstract

Spectrum management has been identified as a crucial step towards enabling the technology of a cognitive radio network (CRN). Most of the current works dealing with spectrum management in the CRN focus on a single task of the problem, e.g., spectrum sensing, spectrum decision, spectrum sharing or spectrum mobility. In this two-part paper, we argue that for certain network configurations, jointly performing several tasks of the spectrum management improves the spectrum efficiency. Specifically, our aim is to study the uplink resource management problem in a CRN where there exist multiple cognitive users (CUs) and access points (APs). The CUs, in order to maximize their uplink transmission rates, have to associate to a suitable AP (spectrum decision), and to share the channels used by this AP with other CUs (spectrum sharing). These tasks are clearly interdependent, and the problem of how they should be carried out efficiently and in a distributed manner is still open in the literature.

In this first part of the paper, we focus on the problem of spectrum sharing in a multi-channel CRN with a single AP. The insight gained from the analysis of this simpler network is used as the building block for analyzing the multiple AP network in the second part of this paper. We formulate the single AP spectrum sharing problem into a non-cooperative power allocation game, in which individual CUs aim at maximizing their transmission rate by finding the suitable power allocation on the available channels. Interestingly, we discover that the set of equilibrium solutions of this game possesses the following optimality properties: 1) any equilibrium solution is the optimum input power allocation scheme in the sense that it maximizes the sum rate of the network if joint decoding at the AP is employed;

M. Hong and A. Garcia are with the Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA. Stephen Wilson is with the Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA.

Part of this manuscript has been accepted by the Proceedings of IEEE INFOCOM 2011 [1].

2) asymptotically, when the number of channels becomes large, any equilibrium solution becomes a Frequency Division Multiple Access (FDMA) strategy, and the maximum system sum rate is achieved without joint decoding. We subsequently propose a set of algorithms for the CUs in the network to achieve such equilibrium solutions in distributed fashion.

I. INTRODUCTION

A. Motivation and Related Work

The problem of distributed spectrum management in the context of CRNs has been under intensive research recently. As pointed out by the authors of [2], spectrum management needs to address four main tasks: 1) *spectrum sensing*, techniques that ensure CUs to find the unused spectrum for communication; 2) *spectrum decision*, protocols that enable the CUs to decide on the best set of channels; 3) *spectrum sharing*, schemes that allow different CUs to share the same set of channels; 4) *spectrum mobility*, rules that require the CUs to leave the channel if licensed users are detected. Many efforts have been devoted to providing solutions to the individual tasks listed above. However, as we will see in this two-part paper, in some CRN scenarios, several of the above tasks become interdependent, and the CUs have to perform these tasks *jointly* to achieve best performance. We thus propose to provide solutions for the joint spectrum decision and spectrum sharing problems in a multi-channel multi-user CRN.

In this two-part paper we focus on investigating an important CRN configuration where such joint spectrum decision and spectrum sharing is desirable. Consider a network with multiple CUs and APs. The APs operate on non-overlapping spectrum bands, and the CUs need to connect to one of the APs for communication. At this stage the CUs essentially perform a spectrum decision task, in which they decide on the best spectrum band to use, i.e., the best AP to connect to. After the AP selection, the CUs can use multiple channels belonging to the associated AP concurrently for transmission, but different CUs interfere with each other if they use the same channel. At this stage the CUs perform a spectrum sharing task, in which multiple CUs use the same spectrum band for communication. This network is a generalization of the single AP network considered in previous literature, e.g., [3], [4] and [5]. It also bears sufficient similarity to the operational model of the IEEE 802.22 cognitive radio standard [6], in which multiple service providers install their respective APs to serve the same geographic region. In the considered network, the CUs face the spectrum decision problem when they select the AP, and they face the spectrum sharing problem when they try to dynamically allocate their communication power across the channels belonging to the selected AP. Clearly, these two problems are strongly interdependent, as on the one hand a particular CU has to select an AP before it can share the spectrum assigned to

this AP with all the other CUs associated with it; on the other hand, after sharing the spectrum, an individual CU may have the incentive to switch to a different AP if it perceives that such action will increase its communication rate. A poor spectrum decision and spectrum sharing scheme will not only lead to unsatisfactory performance for individual CUs, but also result in an unstable system in which CUs are constantly unsatisfied with their current communication rates and consequently changing their AP associations and power allocation indefinitely.

In the first part of this paper, we focus on the spectrum sharing aspect of the above problem. Specifically, we study the uplink power allocation problem in a CRN with a *single* AP. This problem is important in its own right, and the insight gained from studying this problem serves in studying the more complicated network with multiple APs that we discuss in the second part of this paper.

Centralized strategies for resource allocation in multi-carrier single AP network has been extensively studied. In [7] and [8], a joint sub-carrier assignment and power allocation algorithm is proposed for downlink orthogonal frequency division multiplexing (OFDM) network with the objective to optimize utility functions related to throughput and fairness. It has been shown that when the utility function is properly chosen, the optimum access strategy is FDMA, and the downlink throughput can achieve Shannon capacity. In [9], the authors formulate the optimum (in the sense of minimizing the received mean square error) linear transceiver design problem in a multiple access (MAC) intersymbol interference (ISI) channel into an optimum uplink subcarrier allocation and power loading problem, and propose a strongly polynomial algorithm to determine such optimum strategy. [10] proposes a numerical method to compute the capacity of FDMA MAC channel as well as to (near-) optimally assign the channels. The proposed method assigns the channel to different users by solving a (convexly relaxed) optimization problem. [11] and [12] are two recent developments for uplink/downlink resource allocation in multi-carrier systems. The uplink optimality of OFDMA system has been discussed in [13], in which the authors derived sufficient conditions of the channel state as well as the received signal noise ratio for the OFDMA system to achieve maximum uplink system sum rate. We note that the centralized schemes usually assume that the AP carries out all the necessary computations and enforces the resultant optimum policies among the mobile users in the network.

However, such centralized scheme may not be applicable in networks where individual mobile users are autonomous or selfish and have the intention and the ability to deviate from the centralized policies (e.g., in the cognitive radio network). Consequently various distributed algorithms are proposed in the literature, for example, [14], [15] and [4]. In [14], a distributed power allocation scheme is proposed for uplink OFDM systems where the channel state is simplified to having only discrete levels. Notably, this

scheme only requires that each user has the knowledge of *its own* channel state information (CSI), thus the signaling needed for the users to obtain the *global* CSI (as required by the algorithms proposed in, say, [3]) from the AP is greatly reduced. [15] is a recent work considering the uplink dynamic spectrum sharing problem in a multi-carrier multiple service provider cognitive network. The authors developed a distributed algorithm for the users to jointly choose the size of the spectrum as well as the amount of power for transmission. However, in this work the channel is considered to be *flat* for each user, as a result, the users only need to select the *size* of spectrum they need (because any portions of the spectrum of the same size is equivalent to the users), which greatly simplifies the analysis. In [4], the problem of distributed energy-efficient power control in uplink multi-carrier CDMA system is considered. The authors formulate the problem into a game-theoretic framework, and a distributed algorithm is proposed in which each user transmits only on its “best” channel. [16] considered a generalization of a multi-carrier MAC channel, and proposed a distributed iterative water-filling (IWF) algorithm to compute the maximum sum capacity of the system. It is worth mentioning that the algorithms proposed in [4] and [16] both require that the users update their transmission strategies *sequentially*, thus may result in slow convergence when the number of users in the system is large.

In this first part of the paper, we formulate the uplink spectrum sharing problem into a non-cooperative game, in which the CUs in the network try to maximize their individual transmission rate. Due to the structure of the considered multi-channel network, we are able to identify the proposed game as a *potential game* [17], in which the players in the game, although selfish by nature, behave as if they aim at jointly optimizing an objective function (which is called the *potential function*). Such underlying structure of the game allows us to characterize many optimality properties of the equilibrium solution. In particular, the maximum value of the potential function equals to the maximum sum rate achievable for the network. As a result, the CUs can be viewed as jointly optimizing the system sum rate. As far as we know, such interesting relationship between a spectrum sharing game and the system sum rate in multi-carrier single AP network has not been shown in the literature. We then propose three algorithms with convergence guarantees that allow the CUs to reach the equilibrium solution(s) in a distributed fashion. For the sake of rigor, we categorize our algorithms either as “weak convergent”, in which the individual CUs’ strategies converge to *the set* of equilibria, or as “strong convergent”, in which the individual CUs’ strategies converge to an equilibrium point. Such distinction is technical yet necessary, as we will point out in section IV-A, for the reason that unlike most other spectrum sharing games (e.g., [18], [19]), our game generally admits a *connected set of equilibria*, consequently it is possible, at least theoretically, for the algorithm to converge to the set of equilibria without converging to any equilibrium point.

B. Organization of This Work and Notations

This part of the paper is organized as follows. In section II, we present the system under consideration, and formulate the spectrum sharing problem into a non-cooperative game. In section III, we give detailed analysis regarding to the property of the NE of the game. In section IV, we propose distributed algorithms and study their convergence properties. In section V, we show the numerical result. This part of the paper concludes in section VI.

Some notations used in this paper are specified as follows: we use bold lowercase and uppercase letters for vectors and matrices, respectively. The $(i, j)^{th}$ element of a matrix \mathbf{X} is denoted by $[\mathbf{X}]_{i,j}$. For a symmetric matrix \mathbf{X} , $\mathbf{X} \succeq 0$ signifies that \mathbf{X} is positive semidefinite. The trace of a matrix is denoted by $tr(\mathbf{X})$; the determinant of a matrix is denoted by $|\mathbf{X}|$. \mathbf{I}_n is used to denote a $n \times n$ identity matrix. For a vector \mathbf{x} , $diag(\mathbf{x})$ represents a diagonal matrix with its diagonal entries equal to the entries of the vector \mathbf{x} . We use $\rho(\mathbf{X})$ to denote the spectral radius of the matrix \mathbf{X} .

II. PROBLEM FORMULATION

A. System Model

We consider a wireless network with a set $\mathcal{N} \triangleq \{1, 2, \dots, N\}$ CUs and a single AP. Let us normalize the total available bandwidth to 1, and divide it equally into K channels; let the set $\mathcal{K} \triangleq \{1, 2, \dots, K\}$ represents the set of available channels.

The followings are our main assumptions of the network.

- A-1)** The available spectrum can be used exclusively by the CRN, for a relative long period of time.
- A-2)** Each CU can concurrently use all the channels of the AP for transmission, if desired.
- A-3)** The AP is equipped with single-user receivers.

Assumption A-1) can be achieved either under the spectrum property right model in which the licensed networks sell or lease the spectrum to the cognitive network for a period of time for exclusive use, or under the situation that the cognitive network exploits relatively static spectrum white spaces unused by local TV broadcast [20]. Assumption A-3) is congruent with the lack of coordination of the CUs, because an individual CU essentially treats other CUs' transmission as noises. It also allows for implementation of low-complexity receivers at the AP. This assumption is generally accepted in designing distributed algorithm in multiple-access channels [3].

Let $x_i(k)$ denote the complex Gaussian signal transmitted by CU i on channel k ; let $p_i(k) = E[(x_i(k))^2]$ denote the transmitted power of CU i on channel k . Let $z(k) \sim CN(0, n(k))$ denote the white complex Gaussian environment noise experienced at the receiver of AP with mean zero and variance $n(k)$. Let

$h_i(k)$ denote the channel gain coefficient between CU i and the AP on channel k . The signal received at the AP on channel k , denoted by $y(k)$, can then be expressed as:

$$y(k) = \sum_{i=1}^N x_i(k) h_i(k) + z(k). \quad (1)$$

Define $\mathbf{x}_i = [x_i(1), \dots, x_i(K)]^\top$; $\mathbf{z} = [z(1), \dots, z(K)]^\top$; $\mathbf{H}_i = \text{diag}[h_i(1), \dots, h_i(K)]$. Then the received vector signal $\mathbf{y} = [y(1), \dots, y(K)]^\top$ can then be expressed concisely as:

$$\mathbf{y} = \sum_{i=1}^N \mathbf{H}_i \mathbf{x}_i + \mathbf{z}. \quad (2)$$

Let $\mathbf{p}_i = [p_i(1), \dots, p_i(K)]^\top$ be CU i 's transmission power profile; let $\mathbf{p}_{-i} = [\mathbf{p}_1^\top, \dots, \mathbf{p}_{i-1}^\top, \mathbf{p}_{i+1}^\top, \dots, \mathbf{p}_N^\top]^\top$ be the transmission profile of all other CUs except CU i ; let $\mathbf{p} = [\mathbf{p}_1^\top, \dots, \mathbf{p}_N^\top]^\top$ be the system power profile. Define \bar{p}_i to be CU i 's maximum allowable transmission power, then its feasible space can be expressed as $\mathcal{P}_i \triangleq \{\mathbf{p}_i : \mathbf{p}_i \geq \mathbf{0}, \sum_{k=1}^K p_i(k) \leq \bar{p}_i\}$. In this network when single-user receiver is employed at the AP, and when assuming other CUs' power profiles are fixed, the CU i 's maximum achievable rate can be expressed as [21]:

$$R_i(\mathbf{p}_i, \mathbf{p}_{-i}) = \frac{1}{K} \sum_{k=1}^K \log \left(1 + \frac{p_i(k) |h_i(k)|^2}{n(k) + \sum_{j \neq i} p_j(k) |h_j(k)|^2} \right). \quad (3)$$

We assume that each CU $i \in \mathcal{N}$ has the knowledge of its own channel coefficients $\{h_i(k)\}_{k \in \mathcal{K}}$, and the quantity $\sum_{j \neq i} p_j(k) |h_j(k)|^2 + n(k)$ on every channel, which represents the sum of noise plus interference on each channel. This information can be obtained by the AP and fed back to each CU i , as suggested in [4]. We do not assume that individual CU has any information regarding the other CUs' channel coefficients; nor do we require that individual CU i knows the power budget $\{\bar{p}_j\}_{j \neq i}$ or the power allocations \mathbf{p}_{-i} of other CUs.

The problem that the CUs are facing is that under the above sets of system constraints, how should they decide on the policy for efficiently sharing of the available spectrum in a distributed manner? In the following subsection, we formulate such spectrum sharing problem into a game-theoretical framework.

B. A Non-Cooperative Game Formulation

In order to facilitate the development of a distributed algorithm, we model each CU as selfish agent, and its objective is to maximize its own transmission rate. More specifically, when \mathbf{p}_{-i} is fixed, CU i is interested in solving the following optimization problem:

$$\max_{\mathbf{p}_i \in \mathcal{P}_i} \frac{1}{K} \sum_{k=1}^K \log \left(1 + \frac{p_i(k) |h_i(k)|^2}{n(k) + \sum_{j \neq i} p_j(k) |h_j(k)|^2} \right). \quad (4)$$

The solution to this optimization problem is the well-known single-user water-filling solution [21]:

$$\Phi_i^k(\mathbf{p}_{-i}) \triangleq \left[\sigma_i - \frac{n(k) + \sum_{j \neq i} p_j(k) |h_j(k)|^2}{|h_i(k)|^2} \right]^+, \quad \forall k \in \mathcal{K} \quad (5)$$

where $\sigma_i \geq 0$ is the dual variable for the sum-power constraints.

We introduce a non-cooperative spectrum sharing game where the players of the game are the CUs in the network, the utility of each player is its achievable transmission rate, and the strategy of each player is its transmit power profile. We denote this game as $G = \{\mathcal{N}, \mathcal{P}, \{R_i(\cdot)\}_{i \in \mathcal{N}}\}$, where $\mathcal{P} = \prod_{i \in \mathcal{N}} \mathcal{P}_i$ is the joint feasible region of all CUs.

The Nash Equilibrium (NE) of the above game is defined as the strategies $\{\mathbf{p}_i^*\}_{i \in \mathcal{N}}$ satisfying [22]:

$$\mathbf{p}_i^* \in \arg_{\mathbf{p}_i \in \mathcal{P}_i} \max R_i(\mathbf{p}_i, \mathbf{p}_{-i}^*) \quad \forall i \in \mathcal{N}. \quad (6)$$

Intuitively, a NE of the game is a stable point of the system where no player has the incentive to deviate from its current strategy. In the following sections, we will first set out to analyze the properties of the set of NE of game G , and then propose distributed algorithms to reach the set of NE.

III. CHARACTERISTICS AND OPTIMALITY OF THE NE

A. NE as Maximizers of the Potential Function

In order to facilitate the analysis, we introduce the notion of a potential function of the game G . Define a concave function $P : \mathcal{P} \rightarrow \mathbb{R}$:

$$P(\mathbf{p}) \triangleq \frac{1}{K} \sum_{k=1}^K \left(\log \left(n(k) + \sum_{i=1}^N |h_i(k)|^2 p_i(k) \right) - \log(n(k)) \right). \quad (7)$$

We can readily observe that the following identity is true for all $i \in \mathcal{N}$ and $k \in \mathcal{K}$:

$$\frac{\partial P(\mathbf{p})}{\partial p_i(k)} = \frac{\partial R_i(\mathbf{p}_i, \mathbf{p}_{-i})}{p_i(k)}, \quad (8)$$

or similarly, for any \mathbf{p}_i and $\bar{\mathbf{p}}_i \in \mathcal{P}_i$ and for fixed \mathbf{p}_{-i} ,

$$R_i(\mathbf{p}_i, \mathbf{p}_{-i}) - R_i(\bar{\mathbf{p}}_i, \mathbf{p}_{-i}) = P(\mathbf{p}_i, \mathbf{p}_{-i}) - P(\bar{\mathbf{p}}_i, \mathbf{p}_{-i}). \quad (9)$$

We call the function $P(\mathbf{p})$ the *potential function* associated with the game G . Due to the properties (8) and (9), we call the game G a *potential game*. From [17], [23] and [24], we have the following theorem.

Theorem 1: A potential game, say $\mathcal{G} = \{\Omega, \chi, \{U_i\}_{i \in \Omega}\}$, admits at least one pure-strategy NE. If the potential function $P(\cdot)$ associated with the potential game is concave, then a feasible strategy $\mathbf{x}^ \in \chi$ is a NE of the game if and only if it maximizes the potential function, i.e., $\mathbf{x}^* \in \arg_{\mathbf{x} \in \chi} \max P(\mathbf{x})$.*

In light of the above theorem, we immediately have the following Corollary.

Corollary 1: \mathbf{p}^ is a NE of the game G if and only if*

$$\mathbf{p}^* \in \arg \max_{\mathbf{p} \in \mathcal{P}} \frac{1}{K} \sum_{k=1}^K \left(\log \left(n(k) + \sum_{i=1}^N |h_i(k)|^2 p_i(k) \right) - \log(n(k)) \right). \quad (10)$$

B. NE as Optimum Input Strategies that Maximize the Network Sum Rate

It turns out that the potential function (7) has a nice physical interpretation. We show in this subsection that it can be related to the maximum sum rate achievable for the considered network. In order to make the above statement precise, we digress a little to consider the following sum capacity maximization problem of the vector MAC system.

Consider a vector MAC communication system [16] with N users and a single AP. The users and the AP are both equipped with K antennas. Assume the available bandwidth is $\frac{1}{K}$. Let $\hat{\mathbf{x}}_k$ be user k 's transmitting vector signal; let $\hat{\mathbf{H}}_k$ be a $K \times K$ matrix that represents the communication channel between user k and the AP; let $\hat{\mathbf{y}}$ be the aggregated received signal at the AP; let $\hat{\mathbf{z}}$ be the additive Gaussian noise with covariance matrix Σ_z . Then $\hat{\mathbf{y}}$ can be expressed as follows ¹:

$$\hat{\mathbf{y}} = \sum_{i \in \mathcal{N}} \hat{\mathbf{H}}_i \hat{\mathbf{x}}_i + \hat{\mathbf{z}}. \quad (11)$$

Let $\Sigma_i \triangleq E[\hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^T]$ denote user i 's transmission/input covariance. The users are constrained in their individual power output, i.e., the input covariance of the users should satisfy $\text{tr}(\Sigma_i) \leq \bar{p}_i$, $\forall i \in \mathcal{N}$. The capacity-achieving input distribution is known to be a complex Gaussian distribution, and the optimum set of input covariances $\{\Sigma_i^*\}_{i \in \mathcal{N}}$ that maximize the capacity of this system can be found by solving the following problem [16], [21]:

$$\begin{aligned} \max \quad & C(\Sigma) \triangleq \frac{1}{K} \left(\log \left| \sum_{i \in \mathcal{N}} \hat{\mathbf{H}}_i \Sigma_i \hat{\mathbf{H}}_i^T + \Sigma_z \right| - \log |\Sigma_z| \right) \\ \text{s.t.} \quad & \text{tr}(\Sigma_i) \leq \bar{p}_i \quad \forall i \in \mathcal{N} \\ & \Sigma_i \succeq 0 \quad \forall i \in \mathcal{N} \end{aligned} \quad (12)$$

where $\Sigma \triangleq \{\Sigma_i\}_{i \in \mathcal{N}}$ indicates the joint transmission covariance matrix. Notice, that the objective function $C(\Sigma)$ is a concave function [21], hence, it admits a unique maximum value in the feasible region. However, this function is generally not strictly convex, and there are a (connected) set of optimum points that achieve such optimum value.

¹Clearly the received signal $\hat{\mathbf{y}}$ has a similar form with that of the received signal in considered single AP network (cf. (2)), consequently, the results derived in this vector MAC system are instrumental in analyzing our single AP network.

We have the following proposition regarding the input covariance matrices Σ^* that maximizes the sum rate of the system.

Proposition 1: If $\hat{\mathbf{H}}_i$ is diagonal for each i , and if Σ_z is diagonal, then there must exist a set of diagonal input covariance matrices $\{\Sigma_i^\}_{i \in \mathcal{N}}$ that is the optimum solution of the problem (12).*

Proof: We prove this proposition by contradiction. Consider the following convex problem:

$$\begin{aligned} \max \quad & C(\Sigma) \triangleq \frac{1}{K} \left(\log \left| \sum_{i \in \mathcal{N}} \hat{\mathbf{H}}_i \Sigma_i \hat{\mathbf{H}}_i^\top + \Sigma_z \right| - \log |\Sigma_z| \right) \\ \text{s.t.} \quad & \text{tr}(\Sigma_i) \leq \bar{p}_i, \quad \forall i \in \mathcal{N} \\ & \Sigma_i \succeq 0, \quad \forall i \in \mathcal{N} \\ & \Sigma_i \text{ is diagonal}, \quad \forall i \in \mathcal{N}. \end{aligned} \quad (13)$$

Suppose the set of diagonal matrices $\{\Sigma_i^*\}_{i \in \mathcal{N}}$ is an optimum solution of the problem (13), but it is not an optimum solution of the problem (12). For each user $i \in \mathcal{N}$, let $\tilde{\Sigma}_i$ be the (unique) solution of the following optimization problem:

$$\begin{aligned} \max_{\Sigma_i} \quad & \frac{1}{K} \left(\log \left| \hat{\mathbf{H}}_i \Sigma_i \hat{\mathbf{H}}_i^\top + \sum_{j \neq i} \hat{\mathbf{H}}_j \Sigma_j^* \hat{\mathbf{H}}_j^\top + \Sigma_z \right| - \log |\Sigma_z| \right) \\ \text{s.t.} \quad & \text{tr}(\Sigma_i) \leq \bar{p}_i \\ & \Sigma_i \succeq 0. \end{aligned} \quad (14)$$

Then from Theorem 1 of [16], and the assumption that $\{\Sigma_i^*\}_{i \in \mathcal{N}}$ is not optimum solution to the problem (12), there must be at least one user i that, by changing its input covariance matrices from Σ_i^* to $\tilde{\Sigma}_i$, it can strictly improve the objective function, i.e., $\exists i \in \mathcal{N}$, such that $C(\tilde{\Sigma}_i, \Sigma_{-i}^*) > C(\Sigma^*)$.

Let us now find the optimum solution $\tilde{\Sigma}_i$ of the problem (14). Let $\mathbf{N} \triangleq \sum_{j \neq i} \hat{\mathbf{H}}_j \Sigma_j^* \hat{\mathbf{H}}_j^\top + \Sigma_z$, then \mathbf{N} is diagonal because Σ_z , $\{\Sigma_j^*\}_{j \neq i}$ and $\{\hat{\mathbf{H}}_i\}_{i \neq j}$ are all diagonal. Clearly it is also semi-definite. Define a matrix \mathbf{Q} with its elements satisfying:

$$[\mathbf{Q}]_{i,i} = \begin{cases} \frac{1}{\sqrt{[\mathbf{N}]_{i,i}}} & \text{if } [\mathbf{N}]_{i,i} \neq 0 \\ 0 & \text{otherwise} \end{cases}. \quad (15)$$

Then solving problem (14) is equivalent to maximizing the function $\log \left| \mathbf{Q} \hat{\mathbf{H}}_i \Sigma_i \hat{\mathbf{H}}_i^\top \mathbf{Q} + \mathbf{I} \right|$. From Hadamard's inequality [21] and the fact that $\mathbf{Q} \hat{\mathbf{H}}_i$ is diagonal, we have that:

$$\log \left| \mathbf{Q} \hat{\mathbf{H}}_i \Sigma_i \hat{\mathbf{H}}_i^\top \mathbf{Q} + \mathbf{I} \right| \leq \sum_{k \in \mathcal{K}} \log \left(1 + [\hat{\mathbf{H}}_i]_{k,k}^2 [\mathbf{Q}]_{k,k}^2 [\Sigma_i]_{k,k} \right)$$

where the equality is achieved if and only if Σ_i is diagonal, and satisfying $0 \leq \sum_{k=1}^K [\Sigma_i]_{k,k} \leq \bar{p}_i$. Consequently, we conclude that the optimum covariance matrix $\tilde{\Sigma}_i$ of the problem (14) is a diagonal matrix. However, this contradicts the assumption that the set of covariance matrices $\{\Sigma_i^*\}_{i \in \mathcal{N}}$ is an optimum solution of the problem (13), because the new set of covariance matrices $\{\Sigma_{-i}^*, \tilde{\Sigma}_i\}$ is a feasible solution to the problem (13), but $C(\tilde{\Sigma}_i, \Sigma_{-i}^*) > C(\Sigma^*)$. ■

The proof of Proposition 1 points out that when $\{\hat{\mathbf{H}}_i\}_{i \in \mathcal{N}}$ and Σ_z are diagonal, any solution to the optimization problem (13) must be an optimal solution to the original problem (12). In this case, the users only need to select the amount of power on each antenna (power loading transmission scheme) to achieve the maximum sum rate, and the objective function $C(\Sigma)$ of (13) can be reduced to:

$$\begin{aligned} & \frac{1}{K} \log \left| \sum_{i \in \mathcal{N}} \hat{\mathbf{H}}_i \Sigma_i \hat{\mathbf{H}}_i^T + \Sigma_z \right| - \frac{1}{K} \log |\Sigma_z| \\ &= \frac{1}{K} \sum_{k=1}^K \left(\log \left([\Sigma_z]_{k,k} + \sum_{i=1}^N |[\hat{\mathbf{H}}_i]_{k,k}|^2 [\Sigma_i]_{k,k} \right) - \log [\Sigma_z]_{k,k} \right). \end{aligned}$$

Clearly, when the channel matrices and the noise matrix are all diagonal, and the users choose the diagonal transmission strategy, the N -user K -antenna vector MAC channel introduced above is equivalent to the our previously considered N -CU K -channel single AP network (as can be seen from the equivalence of equation (2) and (11)), with the following correspondence of parameters: $\Sigma_i = \text{diag}([p_i(1), \dots, p_i(K)]) \ \forall i \in \mathcal{N}$, $\hat{\mathbf{H}}_i = \text{diag}([h_i(1), \dots, h_i(K)]) \ \forall i \in \mathcal{N}$ and $\Sigma_z = \text{diag}([n(1), \dots, n(K)])$.

The vector MAC capacity maximization problems (12) or (13) can be readily seen as equivalent to the potential function maximization problem (10), and the maximum value of the potential function $P(\mathbf{p})$, say P^* , corresponds to the maximum sum rate achievable for the considered N-CU, K-channel, single AP network. Corollary 1 implies that any NE of the game G maximizes the potential function $P(\mathbf{p})$ among the feasible solutions. Consequently it is also an optimal solution of (13), thus can be viewed as a set of optimal input covariance matrices (or optimum power loading scheme because of the diagonality) that maximizes the achievable sum rate of the system.

However, this result does not imply that the sum of the CUs' rate at a NE achieves the optimal system sum rate. We should point out here that in general, one needs to have both *optimal transmission strategy* and *optimal receiving strategy* to be able to achieve the MAC capacity. In our context, this is to say that in general, at a NE of the game G , although the CUs' transmission strategy is optimal, the sum of individual CUs' rate should be less than the maximum achievable sum rate of the system (notice, that in our considered network, the assumption is that only single-user receiver is implemented at the AP, which is obviously not an optimal receiving strategy).

However, we observe that if a NE of the game G represents a FDMA strategy, then the maximum

system sum rate is achieved using only single-user receiver at the AP. More specifically, if a NE \mathbf{p}^* represents the FDMA strategy, then there is at most a single user $i \in \mathcal{N}$ transmitting on each channel k (i.e., with $\mathbf{p}_i^*(k) > 0$). Define the index set $\mathcal{I}(k) \triangleq \{i : p_i^*(k) > 0\}$, we have that under a FDMA transmission strategy $|\mathcal{I}(k)| \leq 1, \forall k \in \mathcal{K}$. Clearly in this case, the sum rate of the users (when the AP uses single-user receiver) achieves the maximum sum rate of the network:

$$\begin{aligned} \sum_{i \in \mathcal{N}} R_i(\mathbf{p}_i^*, \mathbf{p}_{-i}^*) &= \sum_{i \in \mathcal{N}} \sum_{k=1}^K \log \left(1 + \frac{p_i^*(k) |h_i(k)|^2}{n(k) + \sum_{j \neq i} p_j(k) |h_j(k)|^2} \right) \\ &\stackrel{(a)}{=} \sum_{k=1}^K \sum_{i \in \mathcal{I}(k)} \log \left(1 + \frac{p_i^*(k) |h_i(k)|^2}{n(k)} \right) \\ &\stackrel{(b)}{=} \sum_{k=1}^K \left(\log \left(n(k) + \sum_{i=1}^N |h_i(k)|^2 p_i^*(k) \right) - \log(n(k)) \right) = P(\mathbf{p}^*) \end{aligned} \quad (16)$$

where both (a) and (b) are from the FDMA property of the NE \mathbf{p}^* .

The question remains as in what situation does the NE of the game G represents the FDMA transmission strategy. In the next subsection, we provide an answer to this question by looking at the situation when the available spectrum is arbitrarily finely divided. i.e., $K \gg N$.

C. The Asymptotic Optimality of the NE

In this subsection, we consider the asymptotic situation in which the available bandwidth is arbitrarily finely divided. In this case, the transmission rate for each CU i can be expressed as [21]:

$$R_i(\mathbf{p}_i, \mathbf{p}_{-i}) = \int_0^1 \log \left(1 + \frac{p_i(w) |h_i(w)|^2}{n(w) + \sum_{j \neq i} p_j(w) |h_j(w)|^2} \right) dw \quad (17)$$

where the channel gain $|h_i(w)|^2$ can be viewed as the channel transfer function for CU i to the AP; $n(w)$ becomes the spectral density of the Gaussian noise experienced at the AP on channel k ; \mathbf{p}_i denotes the transmit *power spectral density* of CU i , i.e., $\mathbf{p}_i(w)$ indicates the amount of power CU i transmits on frequency w . In this case, the sum power constraint of each CU should be expressed as: $\int_0^1 p_i(w) dw \leq \bar{p}_i$, and the feasible region \mathcal{P}_i becomes: $\mathcal{P}_i \triangleq \{\mathbf{p}_i : \int_0^1 p_i(w) dw \leq \bar{p}_i, p_i(w) \geq 0 \forall w\}$.

As before, a selfish CU i is interested in solving the following optimization problem:

$$\max_{\mathbf{p}_i \in \mathcal{P}_i} \int_0^1 \log \left(1 + \frac{p_i(w) |h_i(w)|^2}{n(w) + \sum_{j \neq i} p_j(w) |h_j(w)|^2} \right) dw. \quad (18)$$

From the definition of the NE and the solution to the CU's utility maximization problem (18), individual equilibrium transmit spectral \mathbf{p}_i^* must satisfy:

$$p_i^*(w) = \left[\sigma_i - \frac{n(w) + \sum_{j \neq i} p_j^*(w) |h_j(w)|^2}{|h_i(w)|^2} \right]^+, \forall w. \quad (19)$$

We have the following theorem characterizing the system equilibrium transmit spectral $\{\mathbf{p}_i^*\}_{i \in \mathcal{N}}$.

Theorem 2: When the available spectrum is arbitrarily finely divided, and the channel gains are generated according to some continuous distribution, then any NE of the game G represents a FDMA transmission strategy (with probability 1). Moreover, any such NE is efficient, in the sense that the sum of individual users' rates achieves the maximum system sum rate.

Proof: We first show, by contradiction, that any NE represents the FDMA strategy.

Suppose for some channel realization $\{|h_i(w)|^2\}_{i \in \mathcal{N}}$, in the NE of the game a set of CUs $\mathcal{M} \subseteq \mathcal{N}$ are using the frequency w . In another words, we assume the following:

$$p_m^*(w) = \sigma_m - \frac{n(w) + \sum_{j \neq m, j \in \mathcal{M}} p_j^*(w) |h_j(w)|^2}{|h_m(w)|^2} > 0, \quad \forall m \in \mathcal{M}. \quad (20)$$

Then the following is true for all $m \in \mathcal{M}$:

$$n(w) + \sum_{j \in \mathcal{M}} p_j^*(w) |h_j(w)|^2 = \sigma_m |h_m(w)|^2. \quad (21)$$

Thus, for an arbitrary pair of CUs $m_1, m_2 \in \mathcal{M}$: $\sigma_{m_1} |h_{m_1}(w)|^2 = \sigma_{m_2} |h_{m_2}(w)|^2$. However, this equality is satisfied with probability zero (see the proof of Theorem 1 of [3]), because of the fact that σ_{m_1} and σ_{m_2} are constants, and that the channel coefficients are random variables drawn from continuous distributions (Rayleigh distribution or Rician distribution in fading channels). In summary, we claim that the equilibrium transmit power spectral $\{\mathbf{p}_i^*\}$ follows a FDMA scheme with probability 1.

Due to the above FDMA frequency allocation scheme, when $p_i^*(w) > 0$, it must be true that:

$$p_i^*(w) = \left[\sigma_i - \frac{n(w)}{|h_i(w)|^2} \right]^+ = \sigma_i - \frac{n(w)}{|h_i(w)|^2}. \quad (22)$$

Consequently, we can have, for $j \neq i$ (thus $p_j^*(w) = 0$):

$$\begin{aligned} p_j^*(w) &= \left[\sigma_j - \frac{n(w) + |h_i(w)|^2 p_i^*(w)}{|h_j(w)|^2} \right]^+ = 0 \\ \implies \sigma_j - \frac{n(w) + |h_i(w)|^2 p_i^*(w)}{|h_j(w)|^2} &\leq 0 \xrightarrow{(a)} \sigma_j |h_j(w)|^2 \leq \sigma_i |h_i(w)|^2 \end{aligned} \quad (23)$$

where (a) is because of (22). From Theorem 2 of [25], we know that the N -user FDMA scheme maximizes the system sum-rate for a Gaussian multiple access channel if the following is true:

$$\int_0^1 p_i(w) dw = \bar{p}_i, \quad \forall i \in \mathcal{N} \quad (24)$$

$$p_i(w) = \begin{cases} \left[b_i - \frac{n(w)}{|h_i(w)|^2} \right]^+, & \text{if } b_i |h_i(w)|^2 \geq b_j |h_j(w)|^2, \forall j \neq i \\ 0, & \text{otherwise.} \end{cases} \quad (25)$$

Comparing (22)–(23) with (24)–(25), we can readily identify that $\sigma_i = b_i$, and we conclude that $\{\mathbf{p}_i^*\}_{i \in \mathcal{N}}$ achieves the maximum system sum rate. ■

We see from the above derivation that the optimum channel assignment should take into consideration the following three factors [26]: 1) users' channel quality; 2) users' power budget; 3) the noise power. The results derived in Theorem 2 is desirable because in practical multi-carrier systems (e.g. OFDM system), the number of channels is indeed very large compared with the number of users in the system. Consequently, the NE of the spectrum sharing game G represents a desirable outcome in which the CUs in the network share the spectrum efficiently. Interestingly, the authors of [3] has shown that the NE for a uplink power control game represents a time-sharing strategy (which can be viewed as dual to our FDMA strategy), but in a 2-user fading channel system which is very different from the system we consider.

Now the question becomes how such equilibrium point(s) can be reached by individual CU in a distributed fashion. In the next section, we provide three algorithms for such purpose.

IV. THE PROPOSED ALGORITHMS AND CONVERGENCE

From the argument in the previous section, we see that finding the NE of the game G is equivalent to finding $\mathbf{p}^* \in \arg \max_{\mathbf{p} \in \mathcal{P}} \max P(\mathbf{p})$. This is a convex problem and can be solved in a centralized way if all the parameters of the system (e.g., $\{|h_i(k)|^2\}$, $\{\bar{p}_i\}$) are known. However, in a distributed environment, where the CUs are selfish, uncoordinated and not well informed of other CUs' channel coefficients and power budgets, it is not immediately clear how to find such NE point in a distributed fashion.

A. Inapplicability of Conventional IWF Algorithm

We first notice that our model of the network is a special case of a more general network with Gaussian interference channel that has been extensively studied recently, for example, in [18], [19], [27], [28]. In those works, the CUs are transmitter-receiver pairs, and they are interested in allocating their limited transmission power on the set of channels \mathcal{K} to maximize their individual transmission rate. We refer to this network as a Peer-to-Peer (PP) network, while referring to our network as a Access Point (AP) network. In the PP network, we use $|H_{i,j}(k)|^2$ to denote the channel gain from CU i 's transmitter to CU j 's receiver on the k^{th} channel; we use $n_i(k)$ to denote the environmental noise power at the receiver of CU i on channel k . An individual CU i , by transmitting $p_i(k)$ on the k^{th} channel, contributes to every other CU $j \neq i$ in the network the amount of $|H_{i,j}(k)|^2 p_i(k)$ interference at their respective receivers.

Now consider the scenario where all the CUs' receivers are co-located. In this case, for a particular CU i , the set of channel coefficients $\{|H_{i,j}(k)|^2\}_{j \neq i}$ become equal to the value of $|H_{i,i}(k)|^2$; the set of environment noises $\{n_i(k)\}_{i \in \mathcal{N}}$ can be considered equal because the receivers are located at the same place. Consequently, the PP network is equivalent to the AP network.

At this point, we might come to the conclusion that the distributed algorithms developed for the PP network automatically works in the AP network, after all, the PP case is more general than the AP case. However, we show in the following that this is not true. As a matter of fact, the sufficient conditions for the convergence of most algorithms proposed for the PP network are not satisfied in the AP network. As an example, we consider the sufficient condition for the simultaneous IWF algorithm proposed in [27].

Define K nonnegative matrices $\mathbf{H}(k) \in \mathbb{R}_+^{N \times N}$ with their elements defined as follows:

$$[\mathbf{H}]_{q,r}(k) \triangleq \begin{cases} \frac{|H_{r,q}(k)|^2}{|H_{q,q}(k)|^2} & \text{if } r \neq q \\ 0 & \text{otherwise.} \end{cases} \quad (26)$$

Define another nonnegative matrix $\mathbf{H}^{\max} \in \mathbb{R}_+^{N \times N}$ as follows:

$$[\mathbf{H}]_{q,r}^{\max} \triangleq \begin{cases} \max_k \left\{ \frac{|H_{r,q}(k)|^2}{|H_{q,q}(k)|^2} \right\} & \text{if } r \neq q \\ 0 & \text{otherwise.} \end{cases} \quad (27)$$

From Theorem 1 in [27], we have that the simultaneous IWFA algorithm converges to the unique NE of the game if the following is true: $\rho(\mathbf{H}^{\max}) < 1$. In the following, we prove that in AP scenario, this condition can not be satisfied.

From the Perron-Frobenius Theorem [29], we have that there must exist a $N \times 1$ vector $\mathbf{w} > 0$, such that $\|\mathbf{H}^{\max}\|_{\infty}^{\mathbf{w}} = \rho(\mathbf{H}^{\max})$, where $\|\mathbf{A}\|_{\infty}^{\mathbf{w}}$ is the maximum norm of a matrix \mathbf{A} , and is defined as follows:

$$\|\mathbf{A}\|_{\infty}^{\mathbf{w}} \triangleq \max_q \frac{1}{w_q} \sum_{r=1}^Q [\mathbf{A}]_{q,r} w_r, \quad \mathbf{A} \in \mathbf{R}^{Q \times Q}. \quad (28)$$

We next show that in the AP case, there could be no positive vector \mathbf{w} satisfying $\|\mathbf{H}^{\max}\|_{\infty}^{\mathbf{w}} < 1$. Note that we have for all $k \in \mathcal{K}$, $0 \leq \mathbf{H}(k) \leq \mathbf{H}^{\max}$ componentwise, which implies $\|\mathbf{H}(k)\|_{\infty}^{\mathbf{w}} \leq \|\mathbf{H}^{\max}\|_{\infty}^{\mathbf{w}}$ ([29], Chapter 2, Proposition 6.2). Consequently, it is sufficient to prove that there exists $k \in \mathcal{K}$, such that for all $\mathbf{w} > 0$, we must have $\|\mathbf{H}(k)\|_{\infty}^{\mathbf{w}} \geq 1$.

Choose any $k \in \mathcal{K}$. Suppose there exists $\mathbf{w} > 0$ such that $\|\mathbf{H}(k)\|_{\infty}^{\mathbf{w}} < 1$. This implies that: $\max_j \frac{1}{w_j} \sum_{i \neq j} \frac{|H_{i,j}(k)|^2}{|H_{j,j}(k)|^2} w_i < 1$. Then it must be true that for every $j \in \mathcal{N}$, $\frac{1}{w_j} \sum_{i \neq j} \frac{|H_{i,j}(k)|^2}{|H_{j,j}(k)|^2} w_i < 1$, which is equivalent to say that the following N inequalities are true simultaneously:

$$\sum_{i \neq j} |H_{i,j}(k)|^2 w_i < |H_{j,j}(k)|^2 w_j, \quad \forall j \in \mathcal{N}. \quad (29)$$

Recall that when reduced to AP configuration, we have that for all $j \neq i$, $|H_{i,j}(k)|^2 = |H_{i,i}(k)|^2$. Using this equality and adding up N inequalities in (29), we must have:

$$(N-1) \sum_{i \in \mathcal{N}} |H_{i,i}(k)|^2 w_i < \sum_{j \in \mathcal{N}} |H_{j,j}(k)|^2 w_j. \quad (30)$$

Because all the channel coefficients are greater than 0, the above inequality can not be satisfied for any $\mathbf{w} > \mathbf{0}$. Consequently, we prove that there does not exist any $\mathbf{w} > \mathbf{0}$ such that $\|\mathbf{H}(k)\|_\infty^{\mathbf{w}} < 1$. Thus, we must have that $\|\mathbf{H}(k)\|_\infty^{\mathbf{w}} \geq 1$, which in turn says that for all $\mathbf{w} > \mathbf{0}$, we must have $\|\mathbf{H}^{\max}\|_\infty^{\mathbf{w}} \geq 1$, and this implies $\rho(\mathbf{H}^{\max}) \geq 1$. We note further that since $\rho(\mathbf{H}^{\max}) \leq \|\mathbf{H}^{\max}\|$ for any norm (Prop.A.20 in [29]), we must have that $\|\mathbf{H}^{\max}\| \geq 1$ for arbitrary norm. Moreover, we can show similarly that for arbitrary norm, $\|\mathbf{H}(k)\| \geq 1$, $\forall k$, and thus $\rho(\mathbf{H}(k)) \geq 1 \forall k$.

In order to further explain the reason why, in general, algorithms for the PP configuration fail to work in our AP configuration, we observe that almost all the algorithms designed for PP configuration rely on some restrictive conditions of the channel gains to ensure the *uniqueness* of the equilibrium. For example, in [18], the condition $\rho(\mathbf{H}(k)) < 1$, $\forall k \in \mathcal{K}$ ensures the NE of the power allocation game is unique. However as we see in our previous argument, in the AP configuration such condition is not true *for any realization of the channel gains*. As a matter of fact, a straightforward consequence of Corollary 1 is that in general the AP configuration admits *a (connected) set* of equilibrium solutions, as the objective function of the optimization problem (10) is concave, but not strictly concave. A simple example illustrates this point.

Example 1: Consider the network with $N = 2$ CUs, $K = 2$ channels. Let $|h_1(1)|^2 = |h_2(1)|^2 = 1$, $|h_1(2)|^2 = |h_2(2)|^2 = 2$, $n(1) = n(2) = 1$, and let $\bar{p}_1 = \bar{p}_2 = 1$. We can show that both the following two system power profiles $\tilde{\mathbf{p}}$ and $\hat{\mathbf{p}}$ are the NE for the game related to this network:

$$\tilde{p}_1(1) = \frac{3}{4}, \tilde{p}_1(2) = \frac{1}{4}; \tilde{p}_2(1) = 0, \tilde{p}_2(2) = 1; \quad (31)$$

and

$$\hat{p}_1(1) = 0, \hat{p}_1(2) = 1; \hat{p}_2(1) = \frac{3}{4}, \hat{p}_2(2) = \frac{1}{4}. \quad (32)$$

Clearly, from the concavity of the potential function, all the convex combinations of the solutions $\tilde{\mathbf{p}}$ and $\hat{\mathbf{p}}$ also maximize the potential function, hence they are also NEs of the game G .

We conclude the above argument by saying that although the AP network indeed is a special case of the more general PP network, for which distributed algorithms have been developed to reach the NE, these algorithms may not be directly applicable to the AP scenario. Indeed, we will see later in the

simulation section, that by applying the simultaneous IWF algorithm directly to the AP network results in divergence.

B. Proposed Algorithm based on IWF: Weak Convergence

We now proceed to develop algorithms so that the CUs in the single AP network can distributedly reach the NE. In the following we propose two such algorithms.

Algorithm 1: Averaged Iterative-Water Filling Algorithm (A-IWF):

In each iteration t , the CUs do the following.

1) Calculate the best reply power allocation:

$$\Phi_i^k(\mathbf{p}_{-i}^t) \triangleq \left[\frac{1}{\sigma_i} - \frac{n_i(k) + \sum_{j \neq i} |h_j(k)|^2 p_j^t(k)}{|h_i(k)|^2} \right]^+, \forall k \in \mathcal{K} \quad (33)$$

where σ_i ensures $\sum_{k \in \mathcal{K}_w} \Phi_i^k(\mathbf{p}_{-i}^t) = \bar{p}_i$, and let $\Phi_i(\mathbf{p}_{-i}^t) \triangleq [\Phi_i^1(\mathbf{p}_{-i}^t), \dots, \Phi_i^K(\mathbf{p}_{-i}^t)]^\top$.

2) Adjust their power profiles simultaneously according to:

$$\mathbf{p}_i^{t+1} = (1 - \alpha_t) \mathbf{p}_i^t + \alpha_t \Phi_i(\mathbf{p}_{-i}^t) \quad (34)$$

where the sequence $\{\alpha_t\}_{t=1}^\infty$ satisfy $\alpha_t \in (0, 1)$ and :

$$\lim_{T \rightarrow \infty} \sum_{t=1}^T \alpha_t = \infty, \quad \lim_{T \rightarrow \infty} \sum_{t=1}^T \alpha_t^2 < \infty. \quad (35)$$

Algorithm 2: Sequential Iterative-Water Filling Algorithm (S-IWF):

In each iteration t , the CUs adjust their power profiles sequentially² according to:

$$\mathbf{p}_i^{t+1} = \Phi_i(\mathbf{p}_{-i}^t). \quad (36)$$

The convergence properties of the above two algorithms are stated in the following two propositions.

Proposition 2: If all the CUs in the network employ A-IWF algorithm, then their individual power profiles converge to the set of NE of game G .

Proof: Define $\Phi(\mathbf{p}) \triangleq [\Phi_1(\mathbf{p}_{-1}), \dots, \Phi_N(\mathbf{p}_{-N})]^\top$. Define $\mathbf{s}(\mathbf{p}) = \Phi(\mathbf{p}) - \mathbf{p}$. Then from the system point of view the A-IWF algorithm can be written concisely as:

$$\mathbf{p}^{t+1} = (1 - \alpha_t) \mathbf{p}^t + \alpha_t \Phi(\mathbf{p}^t) = \mathbf{p}^t + \alpha_t \mathbf{s}(\mathbf{p}^t). \quad (37)$$

²By “sequential” we mean that the CUs in the set \mathcal{N} take turns in changing their power allocation, and only a single CU gets to act at time t . All other CUs $j \neq i, j \in \mathcal{N}$ keep their power allocation as in time $t - 1$.

We first introduce two lemmas. The proof of Lemma 1 can be found in Appendix A, and we omit the proof of Lemma 2 for brevity.

Lemma 1: There must exist a constant M , with $0 < M < \infty$, such that $\mathbf{s}(\mathbf{p})^\top \nabla_{\mathbf{p}} P(\mathbf{p}) \geq M \|\mathbf{s}(\mathbf{p})\|^2$.

Lemma 2: For two arbitrary vectors $\mathbf{p} \in \mathcal{P}$ and $\bar{\mathbf{p}} \in \mathcal{P}$, and for arbitrary norm $\|\cdot\|$, there must exist two constants $0 < D < \infty$, and $0 < K < \infty$ such that

$$\begin{aligned} \|\mathbf{s}(\mathbf{p}) - \mathbf{s}(\bar{\mathbf{p}})\| &\leq D \|\mathbf{p} - \bar{\mathbf{p}}\| \\ \|\nabla P(\mathbf{p}) - \nabla P(\bar{\mathbf{p}})\| &\leq K \|\mathbf{p} - \bar{\mathbf{p}}\|. \end{aligned} \quad (38)$$

In order to conform to the convention in convex optimization, we define the function $F(\mathbf{p}) = -P(\mathbf{p})$, and we see that $F(\mathbf{p})$ is convex.

Then from the well known Descent Lemma (Lemma 2.1 in [29]), and Lemma 2 we have that:

$$F(\mathbf{p}^{t+1}) \leq F(\mathbf{p}^t) + \alpha_t \mathbf{s}(\mathbf{p}^t)^\top \nabla F(\mathbf{p}^t) + \frac{K}{2} \alpha_t^2 \|\mathbf{s}(\mathbf{p}^t)\|^2 \quad (39)$$

$$\begin{aligned} &\leq F(\mathbf{p}^t) - \alpha_t M \|\mathbf{s}(\mathbf{p}^t)\|^2 + \frac{K}{2} \alpha_t^2 \|\mathbf{s}(\mathbf{p}^t)\|^2 \\ &= F(\mathbf{p}^t) - \alpha_t \|\mathbf{s}(\mathbf{p}^t)\|^2 (M - \alpha_t \frac{K}{2}). \end{aligned} \quad (40)$$

Because α_t goes to 0, then when t large enough, $M - \alpha_t \frac{K}{2} > 0$, and $F(\mathbf{p}^t)$ is monotonically decreasing. Combined with the fact that $F(\mathbf{p}^t)$ is lower bounded, then $\{F(\mathbf{p}^t)\}_{t=1}^\infty$ is a convergent sequence.

From (39), we have that

$$F(\mathbf{p}^{T+1}) \leq F(\mathbf{p}^0) + \sum_{t=0}^T \alpha_t \mathbf{s}(\mathbf{p}^t)^\top \nabla F(\mathbf{p}^t) + \sum_{t=0}^T \frac{K}{2} \alpha_t^2 \|\mathbf{s}(\mathbf{p}^t)\|^2.$$

It is clear that $\|\mathbf{s}(\mathbf{p}^t)\|^2$ is upper bounded, and we have $\lim_{T \rightarrow \infty} \sum_{t=1}^T \alpha_t^2 < \infty$, so $\lim_{T \rightarrow \infty} \sum_{t=0}^T \frac{K}{2} \alpha_t^2 \|\mathbf{s}(\mathbf{p}^t)\|^2 < \infty$. Because $\lim_{T \rightarrow \infty} F(\mathbf{p}^{T+1})$ converges, we must have

$$\begin{aligned} \lim_{T \rightarrow \infty} \sum_{t=0}^T \alpha_t \mathbf{s}(\mathbf{p}^t)^\top \nabla F(\mathbf{p}^t) &> -\infty \\ \lim_{T \rightarrow \infty} \sum_{t=0}^T \alpha_t \mathbf{s}(\mathbf{p}^t)^\top \nabla P(\mathbf{p}^t) &< \infty. \end{aligned} \quad (41)$$

From Lemma 1, we have

$$M \lim_{T \rightarrow \infty} \sum_{t=0}^T \alpha_t \|\mathbf{s}(\mathbf{p}^t)\|^2 \leq \lim_{T \rightarrow \infty} \sum_{t=0}^T \alpha_t \mathbf{s}(\mathbf{p}^t)^\top \nabla P(\mathbf{p}^t) < \infty.$$

Consequently it is clear that we must have $\liminf_{t \rightarrow \infty} \|\mathbf{s}(\mathbf{p}^t)\| = 0$. We show in the following that in fact we have a stronger result that $\lim_{t \rightarrow \infty} \|\mathbf{s}(\mathbf{p}^t)\| = 0$. Suppose not, then $\limsup_{t \rightarrow \infty} \|\mathbf{s}(\mathbf{p}^t)\| > 0$. In

this case there must exist a $\epsilon > 0$ such that the subsequences $\{\tau(n) : \|\mathbf{s}(\mathbf{p}^{\tau(n)})\| < \epsilon, \|\mathbf{s}(\mathbf{p}^{\tau(n)+1})\| \geq \epsilon\}$ and $\{u(n) : \epsilon \leq \|\mathbf{s}(\mathbf{p}^t)\| \leq 2\epsilon, \forall t \in (\tau(n), u(n) - 1), \|\mathbf{s}(\mathbf{p}^{u(n)})\| > 2\epsilon\}$ are both infinite.

For a specific n , the following is true:

$$\begin{aligned} \|\mathbf{s}(\mathbf{p}^{\tau(n)+1})\| - \|\mathbf{s}(\mathbf{p}^{\tau(n)})\| &\leq D\|\mathbf{p}^{\tau(n)+1} - \mathbf{p}^{\tau(n)}\| \\ &\leq D\alpha_{\tau(n)}\|\mathbf{s}(\mathbf{p}^{\tau(n)})\|. \end{aligned} \quad (42)$$

Thus, there exists a N^* such that for all $n > N^*$, we must have $\|\mathbf{s}(\mathbf{p}^{\tau(n)})\| \geq \frac{\epsilon}{2}$.

We also have the following:

$$\begin{aligned} \epsilon &< \|\mathbf{s}(\mathbf{p}^{u(n)})\| - \|\mathbf{s}(\mathbf{p}^{\tau(n)})\| \leq D\|\mathbf{p}^{u(n)} - \mathbf{p}^{\tau(n)}\| \\ &\leq D \sum_{t=\tau(n)}^{t=u(n)-1} \alpha_t \|\mathbf{s}(\mathbf{p}^t)\| \leq D \sum_{t=\tau(n)}^{t=u(n)-1} \alpha_t 2\epsilon \end{aligned} \quad (43)$$

which implies

$$\frac{1}{2D} < \sum_{t=\tau(n)}^{t=u(n)-1} \alpha_t. \quad (44)$$

From our previous derivation, we also have $\lim_{T \rightarrow \infty} \sum_{t=0}^T \alpha_t \|\mathbf{s}(\mathbf{p}^t)\|^2 < \infty$. Then for any $\delta > 0$ there must exists a $\hat{N}(\delta)$ such that for all $n > \hat{N}(\delta)$: $\sum_{t=\tau(n)}^{t=u(n)-1} \alpha_t \|\mathbf{s}(\mathbf{p}^t)\|^2 \leq \delta$.

Take $\delta = \frac{\epsilon^2}{8D}$, and take $n > \max\{N^*, \hat{N}(\frac{\epsilon^2}{8D})\}$, then we have

$$\frac{\epsilon^2}{4} \sum_{t=\tau(n)}^{t=u(n)-1} \alpha_t \leq \sum_{t=\tau(n)}^{t=u(n)-1} \alpha_t \|\mathbf{s}(\mathbf{p}^t)\|^2 \leq \frac{\epsilon^2}{8D} \quad (45)$$

which implies $\sum_{t=\tau(n)}^{t=u(n)-1} \alpha_t \leq \frac{1}{2D}$. This is a contradiction to (44). Thus, we conclude that $\limsup_{t \rightarrow \infty} \|\mathbf{s}(\mathbf{p}^t)\| = 0$, and consequently $\lim_{t \rightarrow \infty} \|\mathbf{s}(\mathbf{p}^t)\| = 0$.

From $\lim_{t \rightarrow \infty} \|\mathbf{s}(\mathbf{p}^t)\| = 0$ we see that the limit point \mathbf{p}_m^* of any converging subsequence of $\{\mathbf{p}^t\}$, say $\{\mathbf{p}^{t_m}\}_{m=1}^\infty$, must satisfy $\Phi(\mathbf{p}_m^*) = \mathbf{p}_m^*$, which is sufficient condition to ensure that \mathbf{p}_m^* is a NE of the game G . Consequently \mathbf{p}_m^* must maximize the function $P(\mathbf{p})$ (from Corollary 1), and this implies that the entire sequence $\{P(\mathbf{p}^t)\}_{t=1}^\infty$ converges to the value $P^* \triangleq \max_{\mathbf{p} \in \mathcal{P}} P(\mathbf{p})$. It also implies that the sequence $\{\mathbf{p}^t\}_{t=1}^\infty$ converge the set of NE of the game G , or in other words, every limit point of $\{\mathbf{p}^t\}_{t=1}^\infty$ is a NE of the game G . ■

Proposition 3: If all the CUs in the network employ S-IWF algorithm, then their individual power profiles converge to the set of NE of game G . Moreover, the potential function $\{P(\mathbf{p}^t)\}_{t=1}^\infty$ is non-decreasing with respect to iteration step t , i.e., $P(\mathbf{p}^{t+1}) \geq P(\mathbf{p}^t)$.

Proof: It is easy to see that this algorithm corresponds to the nonlinear Gauss-Seidel algorithm in solving constrained optimization problem [29], thus the general theory for the convergence of algorithm can be applied (e.g., [29] Proposition 3.9). ■

The S-IWF algorithm is actually a simplification of the algorithm proposed in [16]. We introduce this algorithm here and briefly discuss its convergence analysis because it will be useful in our analysis in the second part of this paper. We need to point out here that the convergence behaviors characterized for A-IWF and S-IWF are *set convergence*, i.e., the distance between the sequence $\{\mathbf{p}^t\}_{t=1}^{\infty}$ and the set of NE decreases to zero. Theoretically, it is possible that multiple limit points exist for such sequence, hence this convergence behavior is weaker than the “strong convergence”, in which the sequence $\{\mathbf{p}^t\}_{t=1}^{\infty}$ admits a single limit point in the set of NE. In practice though, convergence of the sequence $\{\mathbf{p}^t\}_{t=1}^{\infty}$ is always observed³. However, for the sake of rigor, in the next subsection we propose a third algorithm which converges *strongly* to the set of NE.

C. Proposed Algorithm based on Gradient Descent: Strong Convergence

Algorithm 3: Projected Gradient Descent Algorithm:

In each iteration t , the CUs do the following.

- 1) Calculate the gradient of the potential function:

$$\nabla_{\mathbf{p}_i} P(\mathbf{p}^t) = \left[\frac{|h_i(1)|^2}{n(1) + \sum_{j=1}^N |h_j(1)|^2 p_j^t(1)}, \dots, \frac{|h_i(K)|^2}{n(K) + \sum_{j=1}^N |h_j(K)|^2 p_j^t(K)} \right]^\top. \quad (46)$$

- 2) Adjust their power profiles simultaneously according to:

$$\mathbf{p}_i^{t+1} = [\mathbf{p}_i^t + \alpha_t \nabla_{\mathbf{p}_i} P(\mathbf{p}^t)]_{\mathcal{P}_i} \triangleq \Psi_i(\mathbf{p}^t) \quad (47)$$

where the sequence $\{\alpha_t\}_{t=1}^{\infty}$ satisfy $\alpha_t \in (0, 1)$ and (35); the operator $[\cdot]_{\mathcal{P}_i}$ represents the projection on to the space \mathcal{P}_i .

Clearly, this algorithm is based on the classical projected gradient descent algorithm for solving nonlinear optimization problem, but with diminishing stepsize α_t . In order to prove the convergence of this algorithm, we first introduce the notion of *Quasi-Fejér convergence* [30], [31], [32].

Definition 1: A sequence $\{y^t\}$ is *Quasi-Fejér convergent* to a set $U \subseteq \mathbf{R}^n$ if for every $u \in U$ there is a sequence $\{\epsilon_t\}$ such that $\epsilon_t \geq 0$, $\sum_{t=0}^{\infty} \epsilon_t < \infty$ and $\|y^{t+1} - u\|^2 \leq \|y^t - u\|^2 + \epsilon_t$, $\forall t$.

The Quasi-Fejér sequence has the following property [31], [32].

³The S-IWF algorithm proposed in [16] for vector MAC channel also converges to the *set* of optimum points similarly as ours, and in practice it has been observed that such algorithm always converges to a single point.

Theorem 3: If $\{y^t\}$ is Quasi-Fejér convergent to a nonempty set $U \subseteq \mathbf{R}^n$, then $\{y^t\}$ is bounded. If furthermore a limit point y^ of $\{y^t\}$ belongs to U , then $\lim_{k \rightarrow \infty} y^k = y^*$, i.e., the sequence converges to a single point in U .*

Using the notion of Quasi-Fejér convergence, we have the following strong convergence result for Algorithm 3. Please see Appendix B for proof.

Proposition 4: The projected gradient descent algorithm is Quasi-Fejér convergent to the set of NE of game G , with error term $\epsilon_t \triangleq 2\alpha_t (\Psi(\mathbf{p}^t) - \mathbf{p}^t)^\top \nabla_{\mathbf{p}} P(\mathbf{p}^t)$. Moreover, the sequence $\{\mathbf{p}^t\}_{t=1}^\infty$ generated by this algorithm converges to a point in the set of NE.

D. Discussion

We first note that all the three algorithms proposed in the previous subsections can be carried out in a distributed fashion. That is, in order to carry out the computations in each iteration (mainly to compute $\Phi_i(\cdot)$ or $\Psi_i(\cdot)$) of the algorithms, the CUs do not need to know the behavior of other CUs in the network. Instead, an individual CUs i only needs to know the aggregated *interference plus noise* (IPN) contributed by all other CUs on each channel: $IPN_i(k) \triangleq n(k) + \sum_{j \neq i} |h_j(k)|^2 p_j(k)$, $\forall k \in \mathcal{K}$. As suggested by [4], this information can be fed back to the CUs by the AP. In fact, the AP only needs to *broadcast* the quantity $\{n(k) + \sum_{i \in \mathcal{N}} |h_j(k)|^2 p_j(k)\}_{k \in \mathcal{K}}$ to the CUs, and individual CU i can subtract its contribution and calculate $\{IPN_i(k)\}_{k \in \mathcal{K}}$. We can also show that, similarly as in the previous two subsections, that a more general case of the algorithm where each CU i adopts different sequences of update coefficients (say $\{\alpha_t^i\}_{t=1}^\infty$) also converges, as long as each sequence $\{\alpha_t^i\}_{t=1}^\infty$ satisfies the conditions in (35).

As stated previously, the theoretical categorization of the algorithms by their convergence behaviors is necessary, because it is generally not possible for the game G to have a single equilibrium point. Although for the algorithms in both categories, the potential function (or equivalently the sum capacity) converges to the single optimum point, the convergence behavior of the underlying CUs' strategies are more involved. Simply claiming the algorithm to be "convergent" might be too ambiguous and sometimes misleading⁴. We observe that many iterative water-filling based algorithms for calculation of the capacity for vector MAC and broadcast channels, for example the algorithms in [13], [16] and [33], can only be theoretically proven to be weakly convergent (in which the optimum capacity is attained in the limit, but the underlying sequence converges to the optimum set), although in practice they generally converges to a single optimum point.

⁴Indeed, in many situations convergence to a set leads to oscillation of the sequence. For example, the sequence $\{\frac{1}{t} + \sin(0.5\pi t)\}_{t=1}^\infty$ converges to the set $\{-1, 0, 1\}$.

For the descent algorithm, note that if the update step size is a constant, then the algorithm is also weakly convergent (see Prop. 3.4 of [29])⁵. The descent algorithm with diminishing step size is also used in [34] for network utility maximization with feedback uncertainty, and the problem considered is very different from ours. We remark that, *strong convergence* does not imply *fast convergence*. Indeed, although we are able to show that the projected gradient decent algorithm converges strongly (which is theoretically appealing), in practice it tends to converge much slower than A-IWF and S-IWF. As such, in the second part of this paper, we will only choose A-IWF and S-IWF as building blocks for the joint AP selection and power allocation algorithm.

V. SIMULATION RESULTS

In this section, we demonstrate the performance of the proposed algorithm. We have the following general settings for the simulation. We place multiple CUs and the AP randomly in a $10m \times 10m$ area; we let $d_{i,w}$ denote the distance between CU i and AP w , then the channel gains between CU i and AP w . Unless otherwise noted, $\{|h_{i,w}(k)|^2\}_{k \in \mathcal{K}_w}$ are independently drawn from an exponential distribution with mean $\frac{1}{d_{i,w}^2}$ (i.e., $|h_{i,w}(k)|$ is assumed to have Rayleigh distribution).

Fig. 1 shows a typical realization of the three algorithms analyzed in this paper, in a network with 10 CUs and 32 channels. It is seen that the values of the potential function generated by these algorithms converge to the maximum system capacity quickly, but the sum rate of the CUs (hence individual power profiles) converges slowly for the projected gradient descent algorithm.

Fig. 2 partly quantifies the convergence speed of different algorithms. In this figure, we compare the absolute difference between the maximum system sum rate and the values of the potential function generated by different algorithms (i.e., $\{|P(\mathbf{p}^t) - P^*|\}$), in a network with 20 user and 64 channels. We observe that both the A-IWF and S-IWF algorithms converge relatively fast while the projected gradient descent algorithm, as seen in Fig.1, converges slowly. We have also studied the performance of simultaneous IWF algorithm [27], which clearly diverges in our single AP network. Such phenomenon has been partially explained in Section IV-A. In Fig. 3, we characterize the convergence behavior of the sum of the CUs' rate $R(t) \triangleq \sum_{i \in \mathcal{N}} R_i(\mathbf{p}^t)$, by plotting the relative difference between $R(t)$ and $R(100)$: $\frac{|R(100) - R(t)|}{R(100)}$. Such metric can be viewed as related to the convergence speed of the algorithm. We see that for network with 128 channels and with increasing number of CUs, S-IWF converges increasingly

⁵Consequently, most algorithms proposed for potential games based on projected gradient methods (e.g., those in [24]) can also be categorized as weak convergence when the potential function is concave but not strictly concave.

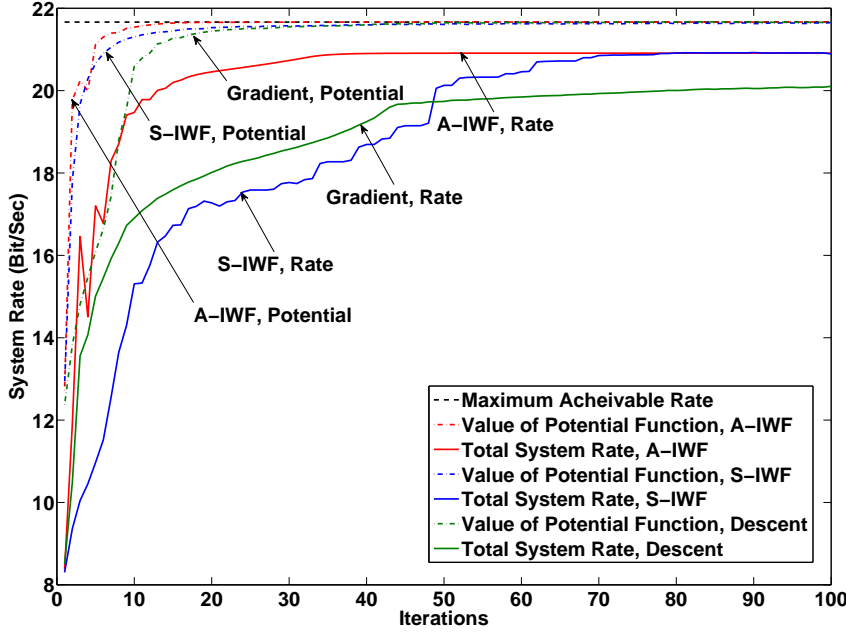


Fig. 1. A particular realization of the algorithms with $K=32$ and $N=10$.

slowly. Such behavior of the S-IWF is intuitively considering the sequential nature of the algorithm. We note that each point in both of these two figures is an average of 100 independent runs of the respective algorithms.

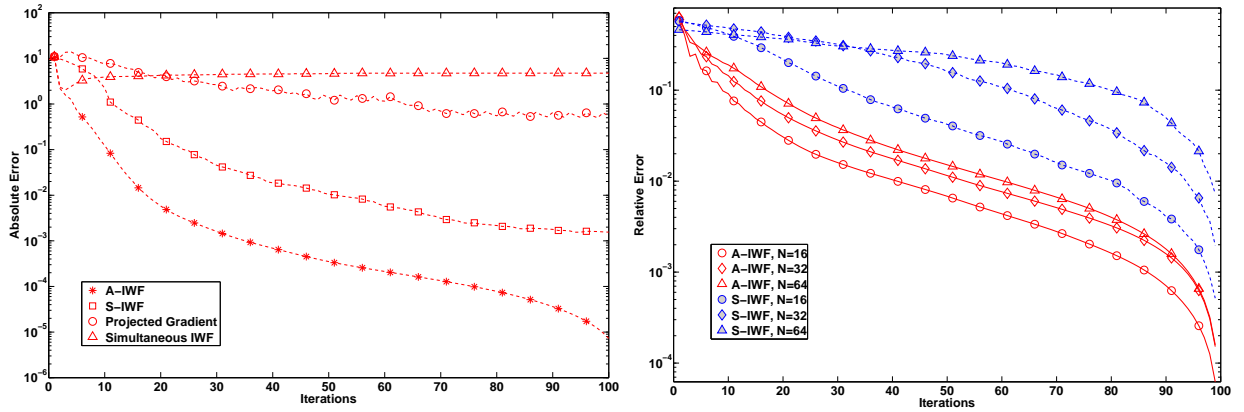


Fig. 2. Averaged absolute difference between the maximum Fig. 3. Convergence behavior of the sum rate of CUs generated system capacity and the value of potential function generated by by A-IWF and S-IWF with $K=128$. different algorithms. $K=32$, $N=10$.

In Section III-C, we have predicted that for a fixed number of CUs, when the number of channels

becomes large, the CUs tend to share the spectrum in a FDMA fashion, and the sum rate of the users approaches the maximum achievable system sum rate. Fig. 4 and Fig. 5 justify these claims. We say that a channel is collided if more than one CUs are using this channel. We say that a (event of) collision occurs if two CUs are using the same channel⁶. In Fig. 4, we plot the relationship between the number of channels in the system and the number of collided channels as well as the total number of collisions. Clearly, as the number of channels becomes large, both of the above quantities decreases. We also observe that when the number of channels becomes large, the number of collided channels tends to be the same as the total number of collisions, a phenomenon which implies that there tend to be no more than two CUs using a collided channel. In Fig. 5, we show the relative difference between the sum rate of the CUs after 200 iteration of the A-IWF algorithm and the maximum sum rate (i.e., $\frac{|R(200) - P^*|}{P^*}$), when the number of channels becomes large. The decreasing of such relative difference is an indication of increased efficiency of the spectrum sharing among the CUs. We note that each point in both of these two figures is again an average of 100 independent runs of the respective algorithms.

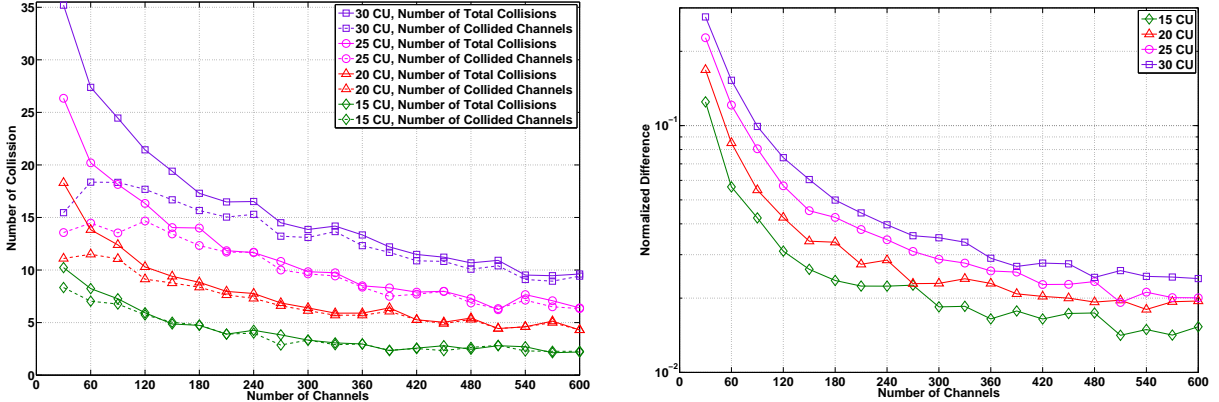


Fig. 4. Averaged number of collisions and averaged number of Fig. 5. Averaged relative difference between the sum rate of collided channels. the CUs and the maximum sum rate of the system.

To quantify the overall efficiency of the spectrum sharing scheme, we plot the normalized system sum rate $\frac{R(200)}{P^*}$ in Fig. 6 for the network with different number of CUs and different number of channels. Clearly the sharing scheme becomes more efficient when the number of channels becomes large. Notice, that in all the previous simulation experiments, we assume that the channel coefficients $\{h_i(k)\}_{k \in \mathcal{K}}$ of a particular CU i to be independent. This is true when the width of each channel is comparable

⁶If $n \geq 2$ CUs are using the same channel, then there are a total number of $\frac{(n-1)(n)}{2}$ collisions occurred.

TABLE I
AVERAGED NUMBER OF COLLISIONS FOR DIFFERENT CHANNEL COHERENT BANDWIDTH B_c WITH $N = 20$.

	Independent	$B_c = 0.1$	$B_c = 0.2$	$B_c = 0.5$	$B_c = 1$
Total Collision, K=300	6.40	7.91	11.12	11.60	14.69
Total Collided Channels, K=300	6.13	7.70	10.06	10.63	13.90
Total Collision, K=600	4.31	5.21	6.00	7.65	12.67
Total Collided Channels, K=600	4.30	5.07	5.81	7.61	12.19

to the coherent bandwidth, denoted as B_c [35]. However, when we divide a fixed spectrum band with arbitrarily large number of channels, the coherent bandwidth eventually becomes larger than the channel width. Indeed, as mentioned in [13], in practice the parallel frequency selective channels are usually correlated. As a result, in Fig .7 we study the spectrum sharing efficiency for a network with $N = 20$ CUs and with networks of different channel coherent bandwidth $B_c = \{1, 0.5, 0.2, 0.1\}$ (recall that our total available bandwidth is normalized to 1). For reference we also plot the case where the channels are assumed to be independent. We observe that large coherent bandwidth reduces the sharing efficiency. This phenomenon can be explained by noticing that when the channel becomes correlated, the event of collision is more likely to happen, as shown in Table I. We again note that each point in both of these two figures and each entry in the table is an average of 100 independent runs of the respective algorithms.

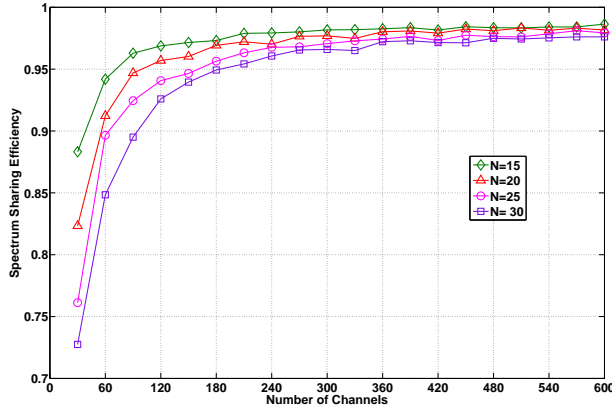


Fig. 6. Comparison of the averaged spectrum sharing efficiency to the number of channels, $N=[15,20,25,30]$.

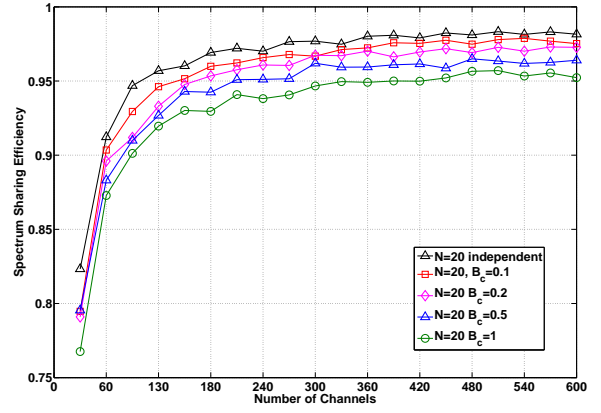


Fig. 7. Comparison of the averaged spectrum sharing efficiencies when the coherent bandwidth is different. $N=20$.

VI. CONCLUSION

In this first part of the paper, we formulate the uplink spectrum sharing problem in a single AP CRN into a non-cooperative game framework. We identify that this game belongs to the family of games called the “potential games”, and we characterize the properties of the proposed game. We then propose three algorithms with different convergence properties that allows the CUs in the network to access the spectrum in a distributed fashion. From simulation we see that the proposed algorithms are able to reach the equilibria of the spectrum sharing game, which represent a set of efficient spectrum sharing strategies.

In the next part of the paper, we will study jointly the spectrum sharing and spectrum decision problem in a CRN with multiple APs. We will see how the algorithms developed in this part of the paper can be used for constructing efficient and distributed joint spectrum decision and spectrum sharing strategies.

APPENDIX A

PROOF OF LEMMA 1

Proof: We first prove Lemma 1. We need to show that the following is true:

$$\sum_{i=1}^N \sum_{k=1}^K \nabla_{p_i(k)} P(\mathbf{p}) s_i^k(\mathbf{p}) \geq \sum_{i=1}^N \sum_{k=1}^K (s_i^k(\mathbf{p}))^2 M \quad (48)$$

where $s_i^k(\mathbf{p}) \triangleq \Phi_i^k(\mathbf{p}_{-i}) - p_i(k)$. It is sufficient to show that for all $i \in \mathcal{N}$, there must exist a constant $0 < M_i < \infty$ such that: $\sum_{k=1}^K \nabla_{p_i(k)} P(\mathbf{p}) s_i^k(\mathbf{p}) \geq \sum_{k=1}^K (s_i^k(\mathbf{p}))^2 M_i$. In the following, we will set out to prove that for all $i \in \mathcal{N}$, there must exist a M_i with $0 < M_i < \infty$, such that:

$$\sum_{k=1}^K (\nabla_{p_i(k)} P(\mathbf{p}) - M_i s_i^k(\mathbf{p})) s_i^k(\mathbf{p}) \geq 0. \quad (49)$$

We notice that

$$\nabla_{p_i(k)} P(\mathbf{p}) = \frac{|h_i(k)|^2}{n(k) + \sum_{j=1}^N |h_j(k)|^2 p_j(k)}, \quad (50)$$

$$s_i^k(\mathbf{p}) = \left[\frac{1}{\sigma_i} - \frac{n(k) + \sum_{j \neq i} |h_j(k)|^2 p_j(k)}{|h_i(k)|^2} \right]_0^{p_{mask}(k)} - p_i(k). \quad (51)$$

We also observe the following equality:

$$\sum_{k=1}^K s_i^k(\mathbf{p}) = \sum_{k=1}^K \Phi_i^k(\mathbf{p}_{-i}) - \sum_{k=1}^K p_i(k) = \bar{p}_i - \bar{p}_i = 0. \quad (52)$$

This can be readily concluded from our previous observation that from each user’s point of view, it is beneficial to allocate all its power for communication. Using (52), we see that in order to prove (49), it is sufficient to prove that for all $k_1 \in \mathcal{K}$ and $k_2 \in \mathcal{K}$ such that

$$s_i^{k_1}(\mathbf{p}) > 0, \text{ and } s_i^{k_2}(\mathbf{p}) < 0 \quad (53)$$

there exists $0 < M_i^{(k_1, k_2)} < \infty$ such that:

$$\nabla_{p_i(k_1)} P(\mathbf{p}) - M_i^{(k_1, k_2)} s_i^{k_1}(\mathbf{p}) \geq \nabla_{p_i(k_2)} P(\mathbf{p}) - M_i^{(k_1, k_2)} s_i^{k_2}(\mathbf{p}) \quad (54)$$

$$\iff \nabla_{p_i(k_1)} P(\mathbf{p}) - \nabla_{p_i(k_2)} P(\mathbf{p}) \geq M_i^{(k_1, k_2)} (s_i^{k_1}(\mathbf{p}) - s_i^{k_2}(\mathbf{p})) > 0. \quad (55)$$

If the above is true, we can take $M_i = \min_{k_1, k_2} M_i^{(k_1, k_2)}$, then for all k_1, k_2 that satisfies (53), we have

$$\begin{aligned} \nabla_{p_i(k_1)} P(\mathbf{p}) - \nabla_{p_i(k_2)} P(\mathbf{p}) &\geq M_i (s_i^{k_1}(\mathbf{p}) - s_i^{k_2}(\mathbf{p})) \\ \implies \nabla_{p_i(k_1)} P(\mathbf{p}) - M_i s_i^{k_1}(\mathbf{p}) &\geq \nabla_{p_i(k_2)} P(\mathbf{p}) - M_i s_i^{k_2}(\mathbf{p}). \end{aligned} \quad (56)$$

Consequently, (49) can be established.

Let us look at the term $s_i^{k_1}(\mathbf{p})$ first. Let us simplify the notation by denoting $s_i^{k_1}(\mathbf{p}) = [A_i^{k_1}]_0^{p_{mask}(k_1)} - p_i(k_1)$, where $A_i^{k_1} \triangleq \frac{1}{\sigma_i} - \frac{n(k_1) + \sum_{j \neq i} |h_j(k_1)|^2 p_j(k_1)}{|h_i(k_1)|^2}$. Because $s_i^{k_1}(\mathbf{p}) > 0$, we must have that $A_i^{k_1} > p_i(k_1) \geq 0$, consequently, we have:

$$0 < [A_i^{k_1}]_0^{p_{mask}(k_1)} - p_i(k_1) \leq A_i^{k_1} - p_i(k_1). \quad (57)$$

We then look at the term $s_i^{k_2}(\mathbf{p})$. We can, similarly as above, also simplify it as $s_i^{k_2}(\mathbf{p}) = [A_i^{k_2}]_0^{p_{mask}(k_2)} - p_i(k_2)$. Because $s_i^{k_2}(\mathbf{p}) < 0$, we must have that $A_i^{k_2} < p_i(k_2) \leq p_{mask}(k_2)$, consequently, we have:

$$0 > [A_i^{k_2}]_0^{p_{mask}(k_2)} - p_i(k_2) \geq A_i^{k_2} - p_i(k_2). \quad (58)$$

As a result of (57) and (58), in order to prove (54), it is sufficient to prove that there exists $0 < M_i^{(k_1, k_2)} < \infty$ such that:

$$\begin{aligned} \nabla_{p_i(k_1)} P(\mathbf{p}) - M_i^{(k_1, k_2)} \times (A_i^{k_1} - p_i(k_1)) \\ \geq \nabla_{p_i(k_2)} P(\mathbf{p}) - M_i^{(k_1, k_2)} \times (A_i^{k_2} - p_i(k_2)). \end{aligned} \quad (59)$$

We see that (59) is equivalent to

$$\begin{aligned} \frac{|h_i(k_1)|^2}{n(k_1) + \sum_{j=1}^N |h_j(k_1)|^2 p_j(k_1)} - M_i^{(k_1, k_2)} \left(\frac{1}{\sigma_i} - \frac{n(k_1) + \sum_{j=1}^N |h_j(k_1)|^2 p_j(k_1)}{|h_i(k_1)|^2} \right) \\ \geq \frac{|h_i(k_2)|^2}{n(k_2) + \sum_{j=1}^N |h_j(k_2)|^2 p_j(k_2)} - M_i^{(k_1, k_2)} \left(\frac{1}{\sigma_i} - \frac{n(k_2) + \sum_{j=1}^N |h_j(k_2)|^2 p_j(k_2)}{|h_i(k_2)|^2} \right). \end{aligned} \quad (60)$$

Define $B_i^{k_1} = \frac{|h_i(k_1)|^2}{n(k_1) + \sum_{j=1}^N |h_j(k_1)|^2 p_j(k_1)}$, and $B_i^{k_2}$ similarly, we have that the above inequality can be simplified to:

$$\begin{aligned} B_i^{k_1} - M_i^{(k_1, k_2)} \left(\frac{1}{\sigma_i} - \frac{1}{B_i^{k_1}} \right) &\geq B_i^{k_2} - M_i^{(k_1, k_2)} \left(\frac{1}{\sigma_i} - \frac{1}{B_i^{k_2}} \right) \\ \iff B_i^{k_1} - B_i^{k_2} &\geq \left(\frac{1}{B_i^{k_2}} - \frac{1}{B_i^{k_1}} \right) M_i^{(k_1, k_2)}. \end{aligned} \quad (61)$$

Now notice that:

$$-\left(A_i^{k_1} - p_i(k_1)\right) + \frac{1}{\sigma_i} = \frac{1}{B_i^{k_1}} \quad \text{and} \quad -\left(A_i^{k_2} - p_i(k_2)\right) + \frac{1}{\sigma_i} = \frac{1}{B_i^{k_2}} \quad (62)$$

and we have from (57) and (58) that

$$\begin{aligned} -\left(A_i^{k_2} - p_i^{k_2}\right) &\geq -s_i^{k_2}(\mathbf{p}) > 0 \\ 0 &> -s_i^{k_1}(\mathbf{p}) \geq -\left(A_i^{k_1} - p_i^{k_1}\right). \end{aligned} \quad (63)$$

We have that $\frac{1}{B_i^{k_1}} < \frac{1}{B_i^{k_2}}$ and $B_i^{k_1} > B_i^{k_2}$. Consequently, (61) is equivalent to

$$M_i^{(k_1, k_2)} \leq \frac{B_i^{k_1} - B_i^{k_2}}{\frac{1}{B_i^{k_2}} - \frac{1}{B_i^{k_1}}} = B_i^{k_2} \times B_i^{k_1}. \quad (64)$$

Now it is clear that we can always find such a $0 < M_i^{(k_1, k_2)} < \infty$, that satisfies the above inequality, because the fact that $B_i^k = \frac{|h_i(k)|^2}{n(k) + \sum_{j=1}^N |h_j(k)|^2 p_j(k)}$ is always bounded above and strictly greater than 0 ($|h_i(k)|^2 > 0 \forall i \in \mathcal{N} \ k \in \mathcal{K}, n(k) > 0 \forall k \in \mathcal{K}$).

Now that we can always find $0 < M_i^{(k_1, k_2)} < \infty$ that satisfies (54), from the argument above, we can see that $0 < M_i = \min_{k_1, k_2 \in \mathcal{K}} M_i^{(k_1, k_2)} < \infty$ must satisfy (49). Thus, take $M = \min_{i \in \mathcal{N}} M_i$, we have that (1) is true. Thus, the proposition is proved. \blacksquare

APPENDIX B

PROOF OF PROPOSITION 4

Proof: The projected gradient algorithm can be written as: $\mathbf{p}^{t+1} = [\mathbf{p}^t + \alpha_t \nabla_{\mathbf{p}} P(\mathbf{p}^t)]_{\mathcal{P}} \triangleq \Psi(\mathbf{p}^t)$, where $\mathcal{P} \triangleq \prod_{i \in \mathcal{N}} \mathcal{P}_i$. We first show that at least one limit point of the sequence \mathbf{p}^* is a NE of the game G . From the Projection Theorem ([29] Sec 3.3 Prop. 3.2) we have that:

$$(\Psi(\mathbf{p}^t) - \mathbf{p}^t - \alpha_t \nabla_{\mathbf{p}} P(\mathbf{p}^t))^{\top} (\mathbf{p}^t - \Psi(\mathbf{p}^t)) \geq 0. \quad (65)$$

Consequently, we have:

$$0 \leq \|\Psi(\mathbf{p}^t) - \mathbf{p}^t\|^2 \leq \alpha_t (\Psi(\mathbf{p}^t) - \mathbf{p}^t)^{\top} \nabla_{\mathbf{p}} P(\mathbf{p}^t). \quad (66)$$

Similarly as in (39), we invoke the descent lemma:

$$\begin{aligned} F(\Psi(\mathbf{p}^t)) &\leq F(\mathbf{p}^t) + (\Psi(\mathbf{p}^t) - \mathbf{p}^t)^{\top} \nabla_{\mathbf{p}} F(\mathbf{p}^t) + \frac{K}{2} \|\Psi(\mathbf{p}^t) - \mathbf{p}^t\|^2 \\ &\stackrel{(a)}{\leq} F(\mathbf{p}^t) + \left(\frac{K}{2} - \frac{1}{\alpha_t}\right) \|\Psi(\mathbf{p}^t) - \mathbf{p}^t\|^2 \\ &\stackrel{(b)}{\leq} F(\mathbf{p}^t) + \left(\frac{K}{2} \alpha_t - 1\right) \alpha_t \|\nabla_{\mathbf{p}} P(\mathbf{p}^t)\|^2 \end{aligned} \quad (67)$$

where (a) is from (66); (b) is because of the non-expansiveness of the projection operator:

$$\begin{aligned} \|\Psi(\mathbf{p}^t) - \mathbf{p}^t\|^2 &= \|\left[\mathbf{p}^t + \alpha_t \nabla_{\mathbf{p}} P(\mathbf{p}^t)\right]_{\mathcal{P}} - [\mathbf{p}^t]_{\mathcal{P}}\|^2 \\ &\leq \alpha_t^2 \|\nabla_{\mathbf{p}} P(\mathbf{p}^t)\|^2. \end{aligned} \quad (68)$$

Thus there must exist a time T^* such that $\forall t > T^*, F(\Psi(\mathbf{p}^t)) \leq F(\mathbf{p}^t)$. From the fact that the function $F(\mathbf{p})$ is lower bounded, we must have that the sequence $F(\mathbf{p}^t)$ converges. An immediate consequence of this result (cf. equation (41)) is that:

$$\sum_{t=1}^{\infty} (\Psi(\mathbf{p}^t) - \mathbf{p}^t)^\top \nabla_{\mathbf{p}} P(\mathbf{p}^t) < \infty. \quad (69)$$

Let \mathbf{p}^* be a limit point of the sequence $\{\mathbf{p}^t\}$, then we must have that $\Psi(\mathbf{p}^*) = \mathbf{p}^*$. This fact combined with the projection theorem implies that for any $\mathbf{y} \in \mathcal{P}$, the following is true:

$$\begin{aligned} 0 &\geq (\mathbf{y} - \Psi(\mathbf{p}^*))^\top (\mathbf{p}^* + \alpha_t \nabla_{\mathbf{p}} P(\mathbf{p}^*) - \Psi(\mathbf{p}^*)) \\ &= \alpha_t (\mathbf{y} - \mathbf{p}^*)^\top \nabla_{\mathbf{p}} P(\mathbf{p}^*). \end{aligned} \quad (70)$$

The last inequality shows that $\mathbf{p}^* \in \arg \max_{\mathbf{p} \in \mathcal{P}} P(\mathbf{p})$, and consequently, \mathbf{p}^* is a NE of the game G .

We then show that the sequence $\{\mathbf{p}^t\}$ is Quasi-Fejér convergent to the set of NE. Using again the Projection Theorem, and (with a little abuse of notation) take \mathbf{p}^* to be *any* NE solution, we have:

$$\begin{aligned} 0 &\leq (\mathbf{p}^* - \Psi(\mathbf{p}^t))^\top (\Psi(\mathbf{p}^t) - \mathbf{p}^t - \alpha_t \nabla_{\mathbf{p}} P(\mathbf{p}^t)) \\ &= (\mathbf{p}^* - \mathbf{p}^t)^\top (\Psi(\mathbf{p}^t) - \mathbf{p}^t - \alpha_t \nabla_{\mathbf{p}} P(\mathbf{p}^t)) \\ &\quad + (\mathbf{p}^t - \Psi(\mathbf{p}^t))^\top (\Psi(\mathbf{p}^t) - \mathbf{p}^t - \alpha_t \nabla_{\mathbf{p}} P(\mathbf{p}^t)). \end{aligned} \quad (71)$$

This is equivalent to:

$$\begin{aligned} &(\mathbf{p}^* - \mathbf{p}^t)^\top (\Psi(\mathbf{p}^t) - \mathbf{p}^t) \\ &\geq \alpha_t (\mathbf{p}^* - \mathbf{p}^t)^\top \nabla_{\mathbf{p}} P(\mathbf{p}^t) + \|\Psi(\mathbf{p}^t) - \mathbf{p}^t\|^2 + \alpha_t (\mathbf{p}^t - \Psi(\mathbf{p}^t))^\top \nabla_{\mathbf{p}} P(\mathbf{p}^t) \\ &\stackrel{(a)}{\geq} \|\Psi(\mathbf{p}^t) - \mathbf{p}^t\|^2 + \alpha_t (\mathbf{p}^t - \Psi(\mathbf{p}^t))^\top \nabla_{\mathbf{p}} P(\mathbf{p}^t) \end{aligned} \quad (72)$$

where (a) is because of the fact that $P(\mathbf{p})$ is concave: $(\mathbf{p}^* - \mathbf{p}^t)^\top \nabla_{\mathbf{p}} P(\mathbf{p}^t) \geq P(\mathbf{p}^*) - P(\mathbf{p}^t) \geq 0$. The distance between \mathbf{p}^* and a arbitrary vector \mathbf{p}^{t+1} can be expressed as follows:

$$\begin{aligned} &\|\mathbf{p}^* - \mathbf{p}^{t+1}\|^2 \\ &= \|\mathbf{p}^* - \mathbf{p}^t\|^2 + \|\mathbf{p}^t - \mathbf{p}^{t+1}\|^2 - 2(\mathbf{p}^* - \mathbf{p}^t)^\top (\mathbf{p}^{t+1} - \mathbf{p}^t) \\ &\stackrel{(a)}{\leq} \|\mathbf{p}^* - \mathbf{p}^t\|^2 - \|\Psi(\mathbf{p}^t) - \mathbf{p}^t\|^2 + 2\alpha_t (\Psi(\mathbf{p}^t) - \mathbf{p}^t)^\top \nabla_{\mathbf{p}} P(\mathbf{p}^t) \\ &\leq \|\mathbf{p}^* - \mathbf{p}^t\|^2 + 2\alpha_t (\Psi(\mathbf{p}^t) - \mathbf{p}^t)^\top \nabla_{\mathbf{p}} P(\mathbf{p}^t) \end{aligned} \quad (73)$$

where (a) is from (72) and the definition of that $\Psi(\mathbf{p}^t) = \mathbf{p}^{t+1}$. Now let us take $\epsilon_t \triangleq 2\alpha_t (\Psi(\mathbf{p}^t) - \mathbf{p}^t)^\top \nabla_{\mathbf{p}} P(\mathbf{p}^t)$. Then we have: $\|\mathbf{p}^* - \mathbf{p}^{t+1}\|^2 \leq \|\mathbf{p}^* - \mathbf{p}^t\|^2 + \epsilon_t$. From (66) and (69) we conclude $\{\epsilon_t\}_{t=1}^\infty$ is non-negative and summable sequence. Because \mathbf{p}^* is an arbitrary NE point, from Definition 1 the sequence $\{\mathbf{p}^t\}$ is Quasi-Fejér convergent to the set of NE of game G . The first part of this proof show that a limit point of $\{\mathbf{p}^t\}$ belongs to the set of NE, consequently, by applying Theorem 3, we see that $\{\mathbf{p}^t\}$ converges to a point in the set of NE. ■

REFERENCES

- [1] M. Hong, A. Garcia, and J. Barrera, "Joint distributed AP selection and power allocation in cognitive radio networks" in *the Proceedings of the IEEE INFOCOM*, 2011, accepted.
- [2] I. F. Akyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Communications Magazine*, pp. 40–48, April 2008.
- [3] L. Lai and H. E. Gamal, "The water-filling game in fading multiple-access channels," *IEEE Transactions on Information Theory*, vol. 54, no. 5, 2008.
- [4] F. Meshkati, M. Chiang, H. V. Poor, and S. C. Schwartz, "A game-theoretic approach to energy-efficient power control in multicarrier CDMA systems," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 1115–1129, 2006.
- [5] M. H. Islam, Y.-C. Liang, and A. T. Hoang, "Joint power control and beamforming for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2415–2419, 2008.
- [6] C. R. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. J. Shellhammer, and W. Caldwell, "IEEE 802.22: the first cognitive radio wireless regional area network standard," *Comm. Mag.*, vol. 47, no. 1, pp. 130–138, 2009.
- [7] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks—part I: Theoretical framework," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 614–624, 2005.
- [8] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks—part II: Algorithm development," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 625–634, 2005.
- [9] Z.-Q. Luo, T. N. Davidson, G. B. Giannakis, and K. M. Wong, "Transceiver optimization for block-based multiple access through ISI channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 4, pp. 1037–1052, 2004.
- [10] W. Yu and J. M. Cioffi, "FDMA capacity of gaussian multiple-access channel with isi," *IEEE Transactions on Communications*, vol. 50, no. 1, pp. 102–111, 2002.
- [11] K. Kim, Y. Han, and S.-L. Kim, "Joint subcarrier and power allocation in uplink OFDMA systems," *IEEE Communication Letters*, vol. 9, pp. 526–528, 2005.
- [12] T. Liu, C. Yang, and L.-L. Yang, "A lower-complexity subcarrier-power allocation scheme for frequency-division multiple-access scheme," *IEEE Transactions on Wireless Communications*, vol. 11, no. 5, pp. 1571–1576, 2010.
- [13] H. Li and H. Liu, "An analysis of uplink OFDM optimality," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 2972–2983, 2007.
- [14] G. He, S. Gault, M. Debbah, and E. Altman, "Distributed power allocation game for uplink ofdm systems," in *Proc. WiOPT*, 2008, pp. 515–521.
- [15] J. Acharya and R. D. Yates, "Dynamic spectrum allocation for uplink users with heterogeneous utilities," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1405–1413, 2009.

- [16] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for gaussian vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 145–152, 2004.
- [17] D. Monderer and L. S. Shapley, "Potential games," *Games and Economics Behaviour*, vol. 14, pp. 124–143, 1996.
- [18] G. Scutari, D. P. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband noncooperative systems based on game theory – part I: Nash equilibria," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, 2008.
- [19] Z.-Q. Luo and J.-S. Pang, "Analysis of iterative waterfilling algorithm for multiuser power control in digital subscriber lines," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–10, 2006.
- [20] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Magazine*, , no. 5, pp. 79–89, 2007.
- [21] T. M. Cover and J. A. Thomas, *Elements of Information Theory, second edition*, Wiley, 2005.
- [22] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*, MIT Press, 1994.
- [23] R. Deb, "A characterization of differentiable potential games," <http://www.econ.yale.edu/rd287/>.
- [24] G. Scutari, S. Barbarossa, and D. P. Palomar, "Potential games: A framework for vector power control problems with coupled constraints," in *the Proceedings of ICASSP 06*, 2006.
- [25] R. Cheng and S. Verdu, "Gaussian multiaccess channels with isi: Capacity region and multiuser water-filling," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 773–785, 1993.
- [26] W. Yu, G. Ginis, and J. M. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 5, pp. 1105–1115, 2002.
- [27] G. Scutari, D. P. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband noncooperative systems based on game theory – part II: Algorithms," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, 2008.
- [28] K. W. Shum, K. K. Leung, and C. W. Sung, "Convergence of iterative waterfilling algorithm for gaussian interference channels," *IEEE Journal on Selected Area in Communications*, vol. 25, pp. 1091–1100, 2007.
- [29] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods, 2nd ed*, Athena Scientific, Belmont, MA, 1997.
- [30] Y. M. Ermoliev, "On the method of generalized stochastic gradient and quasi-fejér sequences," *Cybernetics*, 1969.
- [31] A. N. Iusem, B. F. Svaiter, and M. Teboulle, "Entropy-like proximal methods in convex programming," *Mathematics of Operations Research*, 1994.
- [32] R. Burachik, L. M. G. Drummond, and A. N. Iusem, "Full convergence of the steepest descent method with inexact line search," *Optimization*, pp. 137–146, 1995.
- [33] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna gaussian broadcast channels," *IEEE Transactions on information theory*, vol. 51, no. 4, 2005.
- [34] J. Zhang, D. Zhang, and M. Chiang, "The impact of stochastic noisy feedback on distributed network utility maximization," *IEEE Transactions on Information Theory*, , no. 2, pp. 645–665, 2008.
- [35] A. Goldsmith, *Wireless Communications*, Cambridge University Press, New York, 2005.